|  |
| --- |
| Miguel Ángel Castaño Ibáñez |

# Análisis de Redes Sociales

**Dada la red de “Los Miserables” que indica las conexiones entre los distintos personajes del libro de Victor Hugo (existe conexión si hablan en el libro y el peso de la misma es el número de veces que hablan), y eliminando cada uno de vosotros los nodos (y los arcos relacionados a ellos) que se indican en la tabla adjunta, calcular:**

Datos documento original:

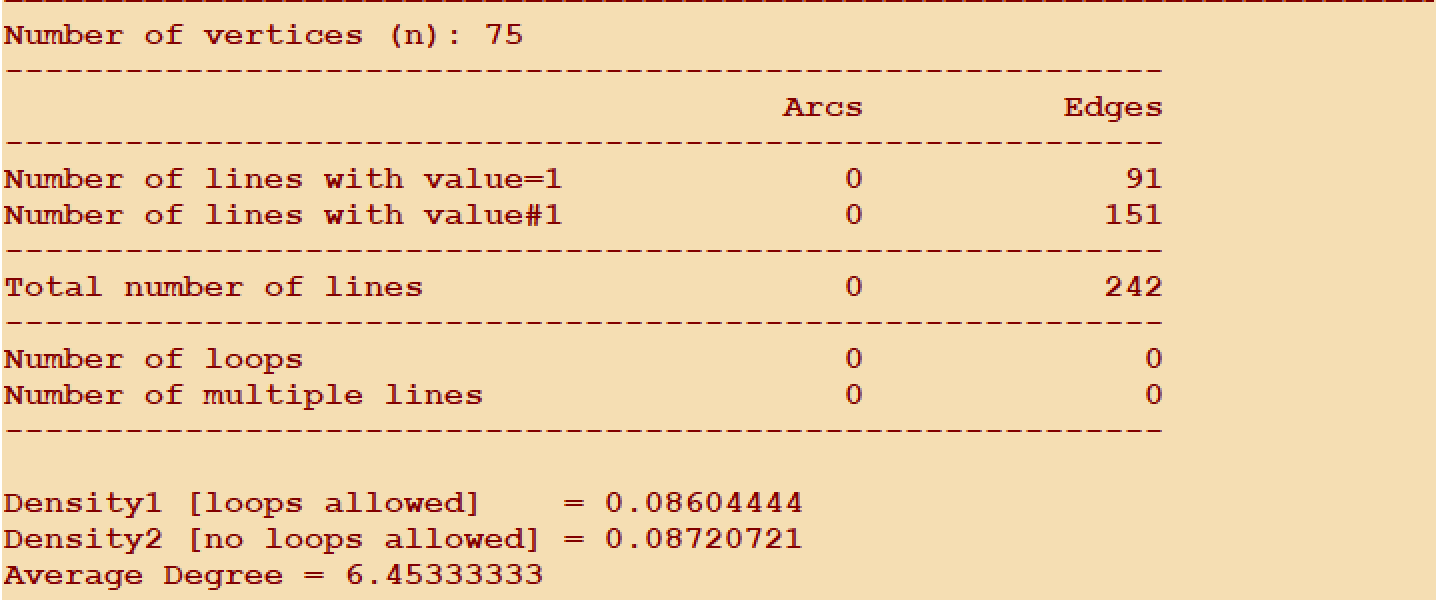
Tabla

Descripción generada automáticamente

A continuación, procederemos a eliminar los nodos que se indica en el ejercicio, en mi caso son 29,30.

Una vez eliminado los 30 nodos finales, obtenemos los siguientes datos de la red ﻿LosMiserablesMaci.net

Datos del documento modificado:



**Diagrama

Descripción generada automáticamente**

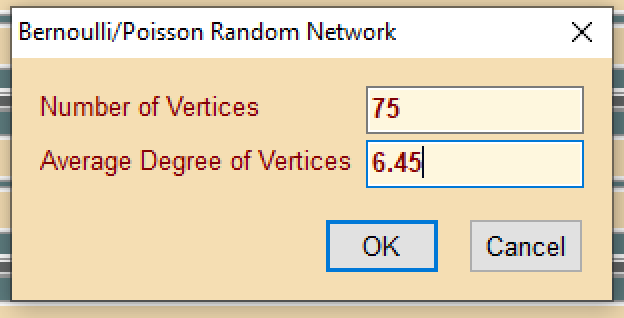
1. **Los parámetros de una red aleatoria con parámetros N y p que tenga un número de nodos y aristas esperadas similar a la red “Los Miserables” e indicar la ruta para hacerlo en Pajek (0.5 puntos).**

Para crear esta red vamos a utilizar el siguiente comando:

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Utilizamos Bermoulli/Poisson en lugar de Total No. of Arcs ya que queremos conseguir una red con un grado medio similar al la original. A continuación, rellenamos los parámetros que se piden a continuación, 75 vértices y el grado medio de 6.45.



Obtenemos la red RandomNetwork.net.

**Imagen que contiene Forma

Descripción generada automáticamente**

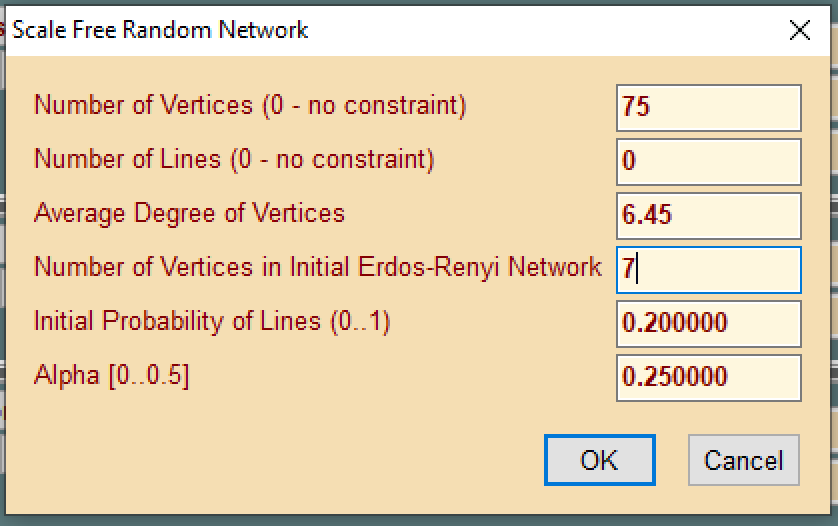
1. **Los parámetros de una red libre de escala que tenga un número de nodos y aristas esperadas similar a la red “Los Miserables” e indicar la ruta para hacerlo en Pajek (0.5 puntos).**

Para crear esta red vamos a utilizar el siguiente comando:

**Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente**

Rellenamos el numero de nodos y el grado medio, también el número de vértices iniciales que en nuestro caso he decidido 7.

****

Obtenemos la red ScaleFree.net.

**Forma

Descripción generada automáticamente**

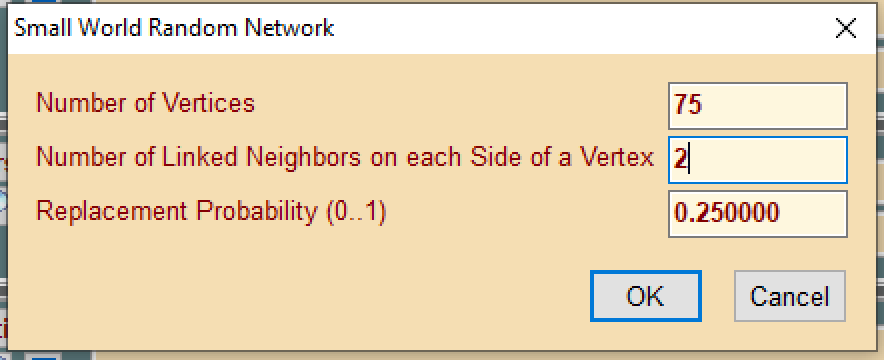
1. **Los parámetros de una red de pequeño mundo que tenga un número de nodos y aristas esperadas similar a la red “Los Miserables” e indicar la ruta para hacerlo en Pajek (0.5 puntos).**

Para crear esta red vamos a utilizar el siguiente comando:

**Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente**

Rellenamos los campos, en primer lugar 75 vértices. Los dos campos restantes indican de probabilidad no muy elevada y no muchos vecinos, para intentar obtener una red de pequeño mundo.

****

Obtenemos la red SmallWorld.net

Forma, Círculo

Descripción generada automáticamente

Tendremos que comprobar si es de pequeño mundo.

**Tabla

Descripción generada automáticamente**

Todos los nodos son conexos, por ello la distancia media de los nodos alcanzables es la misma que la distancia media de la red.

**Texto

Descripción generada automáticamente con confianza media**

Como más o menos se cumple la igualdad podemos afirmar que son de pequeño mundo

1. **Para las cuatro redes (las tres de los apartados anteriores y la original “Los Miserables”), calcular su distribución del grado e indicar si son aleatorias o de libre escala cada una de ellas (1 punto).**

A continuación, vamos a enseñar el gráfico resultante de cada red, donde se muestra la distribución de los grados

Hay que tener en cuenta, para una red aleatoria se obtendrá una distribución normal, se mostrará un histograma con aspecto de distribución gaussiana, mientras que para las redes de libre escala esta distribución tendrá el aspecto similar a una función exponencial decreciente.

Comenzamos viendo la distribución que presenta la red aleatoria creada con el algoritmo Bermoulli/Poisson, donde obtenemos como es normal una distribución gaussiana o dada en una red aleatoria:

**Gráfico, Gráfico de barras

Descripción generada automáticamente**

En la segunda distribución podemos ver la red creada aleatoriamente con el algoritmo red Scale Free (libre de escala), donde obtenemos obviamente una distribución libre de escala:

**Gráfico, Histograma

Descripción generada automáticamente**

El siguiente gráfico representa una red generada aleatoriamente con el algoritmo Small World (pequeño mundo). Como podemos observar esta red también presenta una distribución gaussiana, típica de una red aleatoria.

**Gráfico

Descripción generada automáticamente**

Finalmente vamos a ver nuestra red modificada de los miserables, donde como podemos ver no presenta una distribución normal típicas en las redes aleatorias, si no que más bien podemos apreciar una distribución típica en una red libre de escala, aunque más atenuada que la que creamos con el algoritmo Scale Free (libre de escala) en Pajek. También destaca como hay algunos bastantes nodos con un grado medio más elevado, sin embargo, he decidido considerarla libre de escala por la curva decreciente que presenta.

**Gráfico, Histograma

Descripción generada automáticamente**

1. **Para las cuatro redes (las tres de los apartados anteriores y la original “Los Miserables”), calcular su distancia media e indicar si son de pequeño mundo cada una de ellas (1 punto).**

Podemos decir que una red es de pequeño mundo si se cumple, con más o menos un 20% de error, la igualdad:

* Vamos a comenzar calculando la distancia media para la red aleatoria Bermoulli/Poisson:

Texto

Descripción generada automáticamente con confianza baja

Obtenemos <k>:

Tabla

Descripción generada automáticamente

Una vez hechos los cálculos mostrados en el Excel adjunto “*average\_distance.xls*” obtenemos:

Por lo tanto, podemos concluir que esta red si que la podríamos considerar de pequeño mundo.

* A continuación, calculamos la distancia media para la red aleatoria Scale Free:

Texto

Descripción generada automáticamente

Obtenemos <k>:

Tabla

Descripción generada automáticamente

Hechos los cálculos necesarios en el fichero adjunto “*average\_distance.xls*” obtenemos que:

Por lo tanto, podemos concluir que esta red no es de pequeño mundo

* La siguiente distancia media por calcular va a ser para la red aleatoria Small World:

Imagen que contiene Texto

Descripción generada automáticamente

Obtenemos <k>:

**Tabla

Descripción generada automáticamente**

Este como ya hemos calculado anteriormente si es de pequeño mundo ya que una vez sustituidos los parámetros vemos como se cumple aproximadamente la igualdad:

* Por último, para nuestra red modificada la distancia media va a ser:

Texto

Descripción generada automáticamente con confianza media

Obtenemos <k>

Tabla

Descripción generada automáticamente

Calculados los datos en el fichero adjunto “*average\_distance.xls*” obtenemos que:

Por lo tanto, podemos concluir que esta red si que podría considerarse de pequeño mundo.

1. **Indicar la medida de centralidad que nos parece mejor en este caso y la motivación de esta opinión en esta red (0.5 puntos).**

Creo que en este caso una buena medida de centralidad podría ser la de grado (*degree centrality*), ya que nos podría dar los personajes más influyentes, en base a las conexiones entre estos, es decir un personaje será más importante para la trama conforme tenga más relación con el resto de los personajes. Esta es una de las medidas de centralidad más adecuadas para conectar persona.

Pro otro lado, podemos también tener en cuenta otras medidas de centralidad como ﻿*Weighted All Degree,* es decir, en este caso tenemos en cuenta, además del numero de aristas conectadas de un nodo con otros nodos, el peso de estas conexiones (aristas), que en este caso seria el numero de interacciones entre los personajes.

*Hubs and authorities* también podría ser una buena medida de centralidad, contabilizando más la importancia de las conexiones (aristas) con los nodos más importantes. Considero que esta medida contabilizara sobretodo las conexiones de los personajes principales entre si.

1. **Busca los cinco nodos más importantes en la red y los cinco nodos con una interpretación más diferente en función del criterio de centralidad usado (1.5 puntos).**

Como he dicho en el ejercicio anterior he decidido utilizar como posibles medidas de centralidad: *Degree, Weighted Degree,* y *Hubs and authorities.* Y su posterior análisis en el documento adjunto *“centralitry.xls”*

Comparando estos tres métodos he elegido como medida de centralidad *hubs and authorities* en un (0.7) debido de que a pesar de que los 3 criterios nos dan resultados relativamente parecidos, he querido dar importancia a los nodos hubs, esto se puede traducir en personajes en diferentes tramás en el mismo libro que conectan con personajes de otra trama, por lo tanto, deben de tener una gran importancia en la historia. Por otro lado, también he querido tener en cuenta en un (0.3) *Degree* ya que lo considero un criterio también de importancia. Para ello he tomado como criterio de agregación y dispersión los porcentajes.

* Los 5 nodos más importantes en mi red son(label): **12, 56, 27, 59** y **63**.
* Por otro lado, los nodos que más difieren en función del criterio elegido son: **27, 56, 59, 63, 60**

1. **Para la red “Los Miserables” realizar un clustering por el método de Lovaina e interpretar la modularidad alcanzada. Saca de forma separada cada una de las redes de cada clúster y la red donde todo el clúster se representa como un punto (1.5 puntos).**

Empezamos utilizado el método de Louvain por que damos más importancia a la modularidad, en concreto el método de refinamiento simple, ya que para nuestra red lo considero suficiente

**Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente**

A continuación, encontramos 4 parametros:

* El primer parámetro de resolución que se utiliza para decidir el numero de cortes
* El segundo es la semilla de aleatoriedad, es decir sirve indicarle al algoritmo el numero de permutaciones
* El tercero y el cuarto sirven para fijar el limite de iteraciones en el algoritmo en el primer paso y segundo paso respectivamente

**Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente**

Con los datos que se enseñan en la imagen anterior obtenemos 7 clúster y una modularidad e 0.562 como se muestra a continuación:

**Texto

Descripción generada automáticamente con confianza baja**

Pruebo a cambiar la semilla de alteridad a 10 y vemos que el modularidad y el numero de clúster no ha cambiado. Incluso con 100, que no he decidido mostrar.

**Texto

Descripción generada automáticamente**

A continuación, voy a probar con reducir el parámetro de resolución para intentar cambiar el numero de clúster reduciéndolo a 0.9, 0.8 consiguiendo una modularidad peor, hasta que en 0.7 consigo reducir el numero de clúster hasta 6.

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Por tanto, voy a probar aumentando el parámetro de resolución a 1.1 sin mejorar el modularidad y 1.2 donde consigo aumentar el número de clúster pero sin mejorar la modularidad

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Por lo tanto decido quedarme con la segunda partición donde tengo 7 nodos y una modularidad de 0.563, es decir:

Imagen que contiene tabla, grupo, aire, pequeño

Descripción generada automáticamente

**Texto

Descripción generada automáticamente**

A continuación, extraemos las subredes de cada clúster (estas se encuentran guardada en un fichero de nombre como el color representado en la red original):

* 1º clúster (amarillo) y 2º clúster (verde):

Diagrama, Esquemático

Descripción generada automáticamenteUn dibujo de una persona

Descripción generada automáticamente con confianza baja

* 3º clúster (rojo) y 4º clúster (azul):

Una caricatura de una persona

Descripción generada automáticamente con confianza bajaDiagrama

Descripción generada automáticamente

* 5º clúster (rosa), 6º clúster (blanco) *(este sin representación, ya que es solo un nodo)* y 7º clúster (naranja):

Diagrama, Esquemático

Descripción generada automáticamenteDiagrama

Descripción generada automáticamente

Por último, obtenemos la red donde cada clúster se representa en un solo punto. Voy a poner un mínimo de una relación entre clúster para visualizarlo

Gráfico, Gráfico de líneas

Descripción generada automáticamente

También obtenemos las relaciones entre clúster en la siguiente tabla, sacada del archivo *“cluster\_relations.xls”.*

Numero aristas:

**Imagen que contiene camioneta, camión, escritorio, tabla

Descripción generada automáticamente**

Porcentajes aristas:

Captura de pantalla de un celular con letras

Descripción generada automáticamente

Vemos como los clústeres naranja y verde son los que más relaciones presentan con el resto de los clústeres, pero esto lo analizaremos de forma detallada en el ejercicio siguiente.

1. **Ver la importancia de cada clúster y la capacidad de cada uno de ellos para intermediar (1.5 puntos).**

Para calcular la importancia de cada clúster hay 3 formás:

* **Importancia de nodos en cada clúster.** En le Fichero *“centrality.xls”* ya tenemos la importancia de los nodos calculadas, dividimos en clúster cada nodo y obtenemos el porcentaje de importancia en cada nodo.

Tabla

Descripción generada automáticamente

En este método el clúster 7(naranja) seria el más importante, seguido del clúster 2 (verde). Este método es recomendado para redes pequeñas y no con muchos clústeres por lo que no podríamos utilizarlo para nuestra red, ya que considero que tenemos un numero de clúster elevados respecto al nuero de nodos.

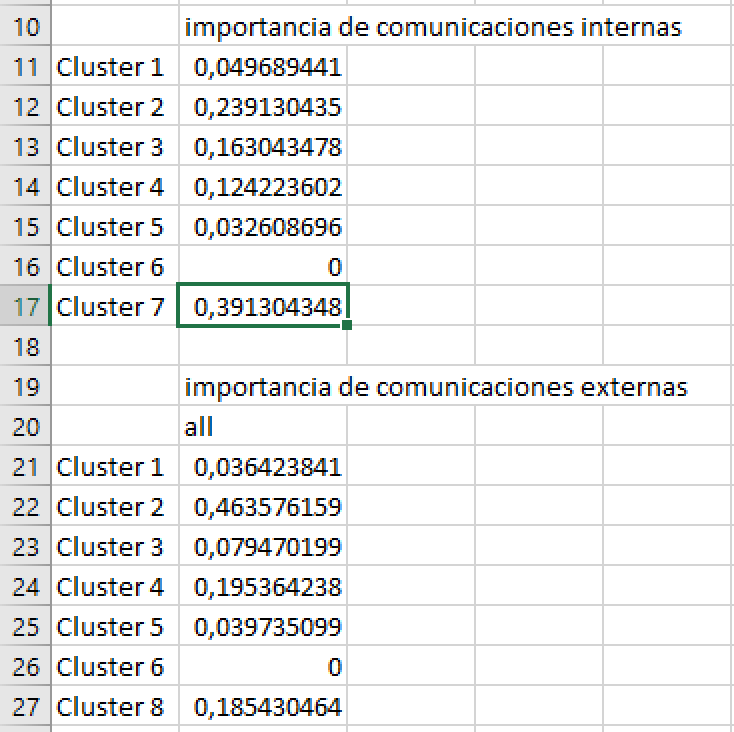
* **Ver el porcentaje de aristas en cada clúster, tanto internas como externas.** En le Fichero *“cluster\_relations.xls”* hemos calculado los porcentajes de aristas internas y externas. Sacado las relaciones de aristas obtenemos los porcentajes.

**Interfaz de usuario gráfica

Descripción generada automáticamente con confianza media**

Visto el numero de aristas de cada clúster obtengo los porcentajes de aristas internas donde el nodo más importante vemos que es el clúster 7 (naranja), seguido por mucha diferencia del clúster 2 (verde).

A continuación, obtenemos el porcentaje de aristas salientes, para ver también la importancia de cada clúster en las comunicaciones con otros clústeres. Aquí vemos como el clúster 2 (verde) obtiene una mayor importancia frente al resto, y como el 7 (naranja) aquí podríamos decir que es de mediana importancia comparándolo también con el 4 clúster (azul).

****

Considero el clúster 2(verde) como el más importante seguido del 7 (naranja). He decidido dar maas importancia al porcentaje de aristas salientes de cada clúster, ya que en mi opinión este clúster 2 (verde) tendrá una mayor centralidad que el 7 (naranja). Esto en el libro se puede traducir en que los personajes del clúster 7 (naranja) aparezcan en una parte concreta y de forma más aislada en la trama del libro, pudiendo ser personajes secundarios o incluso antagonistas, villanos, etc., mientras que los del clúster 2 (verde) aparecerán a lo largo de toda la trama de manera más frecuente, pudiendo ser los protagonistas de la historia

Para mi este método es el más efectivo para medir la importancia de cada clúster ya que lo considero el más apropiado para el tamaño de la red y numero de clústeres, a demás de la información que nos brinda. También tener en cuenta que la información aportada en este método cambia la perspectiva de importancia de clústeres con respecto a los anteriores

* **Importancia de clúster en una red donde cada clúster actúa como un nodo.**

En le Fichero *“centrality\_cluster\_reduced.xls”* encontramos la centralidad de esta red y a continuación tratamos de calcula la importancia de cada nodo donde obtenemos

Texto, Tabla

Descripción generada automáticamente

Para este clúster he tenido más **en cuenta** *Weighted Degree, que Degree,* debido a que puse como condición solo una relación por clúster para crear las aristas. El clúster ganador es el 7(naranja) seguido del 2 (verde). Este método no lo considero tampoco adecuado debido al tamaño de la red, ya que no hay demásiados nodos.

1. **Para los dos nodos indicados para cada uno de vosotros y que sean de distinto clúster mira su centralidad en general en la red (de forma comparativa con el resto de la red), en su clúster y para intermediar entre los distintos clúster (1.5 puntos).**

Vamos a analizar los nodos 31 y 32

Para calcular esta importancia le ha dado 70% par *H&A* y 15% en *Degree* y 15% en *Weighted Degree*

**Cohesividad interna:**

En la red original:

Imagen que contiene Tabla

Descripción generada automáticamente

Imagen que contiene texto, periódico, computadora

Descripción generada automáticamente

Una vez hecho los grupos:

Tabla

Descripción generada automáticamente

Ambos nodos vemos como estarán comidos en la red ya que prácticamente no tienen importancia en la red original y una vez en el clúster tampoco ganan mucha importancia, además conservan sus posiciones relativas 2º y 3º por el final. Quizás podemos decir que el nodo 8 es el que menos cómodo se sentiría de los dos nodos al formar parte del clúster, pero como apenas tiene importancia en la red origina no lo considero relevante.

**Cohesividad externa:**

* **Para el nodo 31.** Al nodo 31 le asignamos un clúster diferente por ejemplo el 8 y aglutinamos todos los clústeres en un solo nodo y vemos como interactúa nuestro nodo con el resto de clúster

Un dibujo de un mapa

Descripción generada automáticamente con confianza baja

Tras hacer el análisis en el fichero *“node\_relation31.xls”* podemos concluir que el nodo 31 solo podría estar relacionado con el clúster 3(rojo) ya que no tiene ninguna relación con el resto de clúster.

En el fichero *“node\_relation31.xls”* se han calculado los porcentajes y las importancias externa y externas

* **Para en nodo 32.** Al nodo 32 le asignamos un clúster diferente el 8 y aglutinamos todos los clústeres en un solo nodo y vemos como interactúa nuestro nodo con el resto de clúster

Imagen de la pantalla de un videojuego

Descripción generada automáticamente con confianza baja

Tras hacer el análisis en el fichero *“node\_relation32.xls”* podemos concluir que el nodo 32 ha sido clasificado en el clúster 3(rojo), pero podría haber sido clasificado también en el clúster 2 (verde), el criterio *Hubs and authorities* ha sido determinante para que este vaya al clúster 2 (rojo) en lugar del 3 (verde) ya que quizás el perteneciente a este ultimo era un nodo no hub

En el fichero *“node\_relation32.xls”* se han calculado los porcentajes y las importancias externa y externas